AF/2714/#

COMBINED TRANSMITTAL OF APPEAL BRIEF TO THE BOARD OF PATENT APPEALS AND INTERFERENCES & PETITION FOR EXTENSION OF TIME UNDER 37 C.F.R. 1.136(a) (Large Entity)	Docket No. SAR 12228

In Re Application Of: **Bergen et al.**

#14 Brief

Serial No.	Filing Date	Examiner	Group Art Unit
08/970,889	11/14/97	M. Padmanabhan	2714

Invention: **METHOD AND APPARATUS FOR EFFICIENTLY REPRESENTING, STORING AND ACCESSING VIDEO INFORMATION**

TO THE ASSISTANT COMMISSIONER FOR PATENTS:

This is a combined Transmittal of Appeal Brief to the Board of Patent Appeals and Interferences and petition under the provisions of 37 CFR 1.136(a) to extend the period for filing an Appeal Brief.

Applicant(s) hereby request(s) an extension of time of (check desired time period):

☐ One month ☐ Two months ☒ Three months ☐ Four months ☐ Five months

from: **June 3, 2000** (*Date*) until: **September 3, 2000** (*Date*)

The fee for the Appeal Brief and Extension of Time has been calculated as shown below:

Fee for Appeal Brief:	$300.00
Fee for Extension of Time:	$870.00
TOTAL FEE FOR APPEAL BRIEF AND EXTENSION OF TIME:	**$1,170.00**

The fee for the Appeal Brief and extension of time is to be paid as follows:

☐ A check in the amount of for the Appeal Brief and extension of time is enclosed.

☒ Please charge Deposit Account No. **20-0782** in the amount of **$1,170.00**
A duplicate copy of this sheet is enclosed.

☒ The Commissioner is hereby authorized to charge payment of the following fees associated with this communication or credit any overpayment to Deposit Account No. **20-0782**
A duplicate copy of this sheet is enclosed.

- ☒ Any additional filing fees required under 37 C.F.R. 1.16.
- ☒ Any patent application processing fees under 37 CFR 1.17.

☒ If an additional extension of time is required, please consider this a petition therefor and charge any additional fees which may be required to Deposit Account No. **20-0782** A duplicate copy of this sheet is enclosed.

09/07/2000 YPOLITE1 00000027 200782 08970889
02 FC:117 870.00 CH

COMBINED TRANSMITTAL OF APPEAL BRIEF TO THE BOARD OF PATENT APPEALS AND INTERFERENCES & PETITION FOR EXTENSION OF TIME UNDER 37 C.F.R. 1.136(a) (Large Entity)	Docket No. SAR 12228

In Re Application Of: **Bergen et al.**

Serial No.	Filing Date	Examiner	Group Art Unit
08/970,889	11/14/97	M. Padmanabhan	2714

Invention: **METHOD AND APPARATUS FOR EFFICIENTLY REPRESENTING, STORING AND ACCESSING VIDEO INFORMATION**

TO THE ASSISTANT COMMISSIONER FOR PATENTS:

This combined Transmittal of Appeal Brief to the Board of Patent Appeals and Interferences and petition for extension of time under 37 CFR 1.136(a) is respectfully submitted by the undersigned:

E J Wall

Signature

Dated: 8/28/00

Eamon J. Wall, Reg. No. 39,414
Thomason Moser & Patterson LLP
595 Shrewsbury Avenue, 1st Floor
Shrewsbury, New Jersey 07702
(732) 530-9404

Certificate of Transmission by Facsimile*

I certify that this document and authorization to charge deposit account is being facsimile transmitted to the United States Patent and Trademark Office (Fax. No.) on (Date)

Signature

Typed or Printed Name of Person Signing Certificate

***This certificate may only be used if paying by deposit account.**

Certificate of Mailing

I certify that this document and fee is being deposited on 8/28/00 with the U.S. Postal Service as first class mail under 37 C.F.R. 1.8 and is addressed to the Assistant Commissioner for Patents, Washington, D.C. 20231. BOX-AF

Laura E Crater

Signature of Person Mailing Correspondence

LAURA E CRATER

Typed or Printed Name of Person Mailing Correspondence

cc: **Sarnoff Corporation**

IN THE UNITED STATES
PATENT AND TRADEMARK OFFICE

PATENT APPLICATION

Appellant(s)	:	Bergen, et al.
Case:		SAR 12228
Serial No.:		08/970,889 Filed: November 14, 1997
Group Art Unit:		2714
Examiner:		Padmanabhan, M.
Title:		METHOD AND APPARATUS FOR EFFICIENTLY REPRESENTING, STORING AND ACCESSING VIDEO INFORMATION

CERTIFICATE OF MAILING under 37 CFR § 1.8(a)

I hereby certify that this correspondence is being deposited on 8/28/00 with the United States Postal Service as first class mail, with sufficient postage, in an envelope addressed to the Assistant Commissioner for Patents, Box AF, Washington, D.C. 20231.

Laura L Crates — Signature

8/28/00 — Date of signature

ASSISTANT COMMISSIONER FOR PATENTS
Box AF
Washington, D. C. 20231

S I R:

BRIEF ON APPEAL

The following appeal brief is submitted pursuant to the Notice of Appeal filed on March 28, 2000 and received by the Patent Office on April 3, 2000 in the above-identified application.

REAL PARTY IN INTEREST

The real party in interest is Sarnoff Corporation.

RELATED APPEALS AND INTERFERENCES

No other appeals or interferences that directly affect, or are directly affected by, or have a bearing on the Board's

decision in the pending appeal, are known to the Appellants' Appellants' legal counsel, or the assignee.

STATUS OF THE CLAIMS

Claims 1-11, 13-14 and 17-26 stand under final rejection, from which rejection this appeal is taken.

STATUS OF AMENDMENTS

The first amendment was filed on October 8, 1999 in response to a first Office Action dated July 12, 1999 (Paper No. 6). In the Office Action, the Examiner noted that claims 1-21 were pending in the application and that claims 1-21 were rejected. In this amendment, claims 12, 15 and 16 were canceled, claims 1-5, 9, 13-14, 17 and 21 were amended, claims 6-8, 10-11 and 18-20 continue unamended and new claims 22-26 were added.

A second amendment was filed on February 24, 2000 in response to a second (Final) Office Action dated January 26, 2000 (Paper No. 8). In the Office Action, the Examiner noted that claims 1-11, 13-14 and 17-26 were pending in the application and rejected. In the response, there were no amendments made. It is noted that the response was filed after a telephone interview with the Examiner on February 22, 2000, which was referenced by the Examiner in an interview summary dated February 25, 2000 (Paper No. 9).

The Examiner responded to Appellants' response of February 24, 2000 with an Advisory Action mailed on March 13, 2000 (Paper No. 11). The Advisory Action reiterated the Examiner's previous position, including his position with respect to the specific teachings of two references.

On March 28, 2000, the Appellants filed a Notice of Appeal from the Examiner's Final Office Action.

SUMMARY OF INVENTION

The present invention is a method and apparatus for comprehensively representing video information in a manner facilitating indexing of the video information. The invention also contemplates an information database suitable for providing scene-based video information to a user.

The process of constructing these scene-based video representation may be conceptualized as a plurality of analysis steps operative upon the appropriate portions of an evolving scene representation. That is, each of a plurality of video processing techniques are employed to operate on at least respective portions of information associated with a particular scene.

The invention comprises the selective use of the following video processing steps to provide a comprehensive method of representing video information in a manner facilitating indexing of the video information: (a) scene segmentation including "key frame" designation; (b) mosaic construction; (c) motion analysis;(d) appearance analysis and (e) ancillary data capture.

Segmentation comprises the process of segmenting a continuous video stream into a plurality of segments, or scenes, where each scene comprises a plurality of frames, one of which is designated a "key frame." The "key frame" may comprise a merged background layer derived from other frames within the scene. Other frames may be represented in terms of differences from the key frame.

Mosaic construction comprises the process of computing, for a given scene or video segment, a variety of "mosaic" representations and associated frame coordinate transforms, such

as background mosaics, synopsis mosaics, depth layers, parallax maps, frame-mosaic transforms, and frame-reference image coordinate transforms. For example, a mosaic of background layers may be used to provide the "key frame" of a scene.

Motion analysis comprises the process of computing, for a given scene or video segment, a description of the scene or video segment. Motion analysis leads to the creation of the associated mosaic representation for the foreground, background and other layers in a scene or segment. Motion analysis may be performed in terms of:

(1) layers of motion and structure corresponding to objects, surfaces and structures at different depths and orientations;
(2) independently moving objects;
(3) foreground and background layer representations; and
(4) parametric and parallax/depth representations for layers, object trajectories and camera motion.

Appearance analysis comprises the process of computing, for a frame or a layer (e.g., background, depth) of a scene or video segment, content-related or appearance attribute information such as color or texture descriptors which are represented as a collection of feature vectors.

Ancillary data capture comprises the process of capturing, through ancillary data streams (time, sensor data, telemetry and the like) or manual entry, ancillary data related to some or all of the scenes or video segments.

For the Board of Patent Appeals & Interferences, appellants' claim 1 (one of the broadest independent claims) is presented below in claim format with elements read on the various figures of the drawings as suggested in M.P.E.P. 1206. Claim 1 positively recites (with reference numerals added):

"A method for comprehensively representing video information in a manner facilitating indexing of the video information, comprising the step of:

segmenting (122) a video stream into a plurality of scenes (710), each of said scenes (S) comprising at least one video frame (F);

dividing (310), using intra-scene motion analysis, at least one of said plurality of scenes into at least one scene foreground layer and a scene background layer;

representing (315) each scene background layer as a mosaic, said background layer mosaic defining a key frame (760) of a respective scene; and

representing (315) each (770, 780) of said at least one video frames forming said scene as a difference between initial video frame imagery (730, 740) and a respective portion of said key frame (760)."

ISSUES

A. Whether Claims 1-3, 11 and 21-23 Are Patentable Under 35 U.S.C. §103(a) over Adelson (U.S. Patent No. 5,706,417 issued January 6, 1999) in view of Yeo et al. (U.S. Patent No. 5,821,945, issued October 13, 1998) and Shibata et al. (Content-Based Structuring of Video Information, 0-8186-7436-9/96, 1996 I.E.E.E.).

B. Whether Claims 4 and 24 are patentable Under 35 U.S.C. §103(a) Over Adelson (U.S. Patent No. 5,706,417 issued January 6, 1999) in view of Yeo et al. (U.S. Patent No. 5,821,945, issued October 13, 1998) and Shibata et al. (Content-Based Structuring of Video Information, 0-8186-7436-9/96, 1996 I.E.E.E.) as applied to Claims 1 and 22, respectively, and further in view of Jaillon et al. (Image Mosaicing Applied To Three-Dimensional Surfaces: 1051-4651/94-1994 I.E.E.E.).

C. Whether Claims 5-8 are Patentable Over Adelson (U.S. Patent No. 5,706,417, issued January 6, 1999), Yeo et al. ((U.S. Patent No. 5,821,945, issued October 13, 1998) and Shibata et al. (Content-Based Structuring of Video Information, 0-8186-7436-9/96, 1996 I.E.E.E.) as applied to Claim 2 and further in view of Jaillon et al. (Image Mosaicing Applied To Three-Dimensional Surfaces: 1051-4651/94-1994 I.E.E.E).

D. Whether Claims 9 and 10 are patentable under 35 U.S.C. §103(a) over Adelson (U.S. Patent No. 5,706,417, issued January 6,

1999) in View of Yeo et al. (U.S. Patent No. 5,821,945, issued October 13, 1998) and Shibata et al. (Content-Based Structuring of Video Information, 0-8186-7436-9/96, 1996 I.E.E.E.) as applied to Claim 2 and further in view of Barber et al. (U.S. Patent No. 5,751,286, issued May 12, 1998).

E. Whether Claims 13-14 are patentable over Adelson (U.S. Patent No. 5,706,417, issued January 6, 1999) in view of Yeo et al. (U.S. Patent No. 5,821,945, issued October 13, 1998) and Shibata et al. (Content-Based Structuring of Video Information, 0-8186-7436-9/96, 1996 I.E.E.E.) as Applied to Claim 1,and further in view of Zhang (U.S. Patent No. 5,635,982, issued June 3, 1997).

F. Whether Claims 17-20 are patentable under 35 U.S.C. §103(a) over Barber et al. (U.S. Patent No. 5,751,286, issued May 12, 1998) in view of Yeo et al. (U.S. Patent No. 5,821,945, issued October 13, 1998) in view of Shibata et al. (Content-Based Structuring of Video Information, 0-8186-7436-9/96, 1996 I.E.E.E.).

G. Whether Claims 25-26 are patentable over Adelson (U.S. Patent No. 5,706,417, issued January 6, 1999) in view of Yeo et al. (U.S. Patent No. 5,821,945, issued October 13, 1998) and Shibata et al. (Content-Based Structuring of Video Information, 0-8186-7436-9/96, 1996 I.E.E.E.) as applied to Claim 22 and further in view of Jaillon et al. (Image Mosaicing Applied To Three-Dimensional Surfaces: 1051-4651/94-1994 I.E.E.E).

GROUPING OF CLAIMS

The rejected claims have been grouped together in the rejection. Appellants urge that each of the rejected claims stands on its own recitation, the claims being considered to be separately patentable for the reasons set forth in more detail infra. The following references are relied on by the Examiner:

Author	**Publication Title/ Reference No.**	**Publication Date**
Adelson	5,706,417	1/6/99
Barber et al.	5,751,286	5/12/98

Jaillon et al.	Image Mosaicing Applied to Three-Dimensional Surfaces, 0-8186-7436-9/96	1994
Shibata et al.	Content-Based Structuring of Video Information, 1051-4651/94	1996
Yeo et al.	5,821,945	10/13/98
Zhang	5,635,982	6/3/97

BRIEF DESCRIPTION OF THE REFERENCES

Adelson discloses a method and apparatus of layered representation for image coding wherein each object, set of objects, or portion of an object in the image having a motion vector significantly different from any other object in the image may be represented by a unique layer. Adelson teaches the representation of an image as a series of N layers ordered by "depth" in an image, where each layer comprises a series of data maps. Standard maps include an intensity map, an attenuation map, a velocity map, and a delta map. Optional maps include a contrast change map, a blur map, a depth map and a surface orientation map. Each map comprises a set of data for discrete two-dimensional locations and, optionally, a time dimension. There is no teaching of the use of a third dimension other than the depth associated with each of the end layers of an image.

Barber discloses an image query system and method wherein images in an image data base are searched in response to queries which include the visual characteristics of the images such as colors, textures, shapes, and sizes as well as by textual tags appended to the images.

Jaillon discloses a method of mosaicing still images lying on a three-dimensional surface. Using a rough model of the three dimensional surface and the parameters of the projection of a still image on that surface, the three dimensional model is "flattened" and the resulting two-dimensional images are merged

using a two-dimensional mosaicing technique. Various corrections are applied on the original still images for Laplacian images and the resulting corrected two-dimensional still image is mapped on an approximate elevational model to allow three-dimensional visualization of the still images. The Jaillon reference finds use in the areas of microscopy and satellite imagery, for example.

Shibata teaches content-based structuring of video information using textual descriptions. In fact, Shibata has absolutely nothing to do with the present invention. Shibata defines (per Section 3.1) "video structuring" as an operation which divides a video sequence into "segments" and describes the hierarchical relations between them. It is also noted that the description in Shibata of the relations between segments is a textual description intended to provide a human readable description of the underlying video scene such that the underlying video may be manually processed by a director or editor within the context of a video editing environment or studio environment, e.g. by a director. Specifically, a descriptive component (DC) is defined by Shibata as key words or elemental words that constitute short sentences which may be divided into several groups (see Section 2.). With respect to video structuring, the categories of visual objects, actions of the object, and state of the object are used. The descriptive components (DCs) are mapped (see FIG. 1) as a script which indicates the presence or absence of particular descriptive components within the video sequence in time.

The "vector expressions" of Shibata are not motion vectors. Rather (per section 3.1), the Shibata "vector expressions" are merely representations of the duration of descriptive components in terms of time or segment length. The Shibata "vector expressions" should not be equated with the motion vectors discussed in the instant patent application. It can be seen in FIG. 2 of Shibata that each "layer" is formed by averaging "basic segments" of a lower layer. That is, as depicted in FIG. 2, where M basic segments are provided, the M^{th} layer includes the M "basic segments." By averaging the vector expressions of respective

adjacent basic segments within the M^{th} layer, an M^{th} minus 1 layer is formed which includes M/2 basic segments. Each of the M/2 basic segments comprises the averaged vector expressions of the two basic segments within the M^{th} layer. Similarly, for each succeeding layer, respective pairs of basic segments or derived (i.e., averaged) basic segments are themselves averaged to produce the next layer. A top layer or M = 1 layer comprises the average of all of the vector expressions of the basic segments forming the M^{th} layer.

Yeo discloses a method for video browsing based on content and structure. The Yeo method arranges video information such that a human browsing through the arranged video information may easily find desired video imagery. Referring to FIG. 1 of the Yeo patent, scene change detection is employed to divide a video screen into a plurality of video "shots," which are then arranged into a plurality of "clusters," where each cluster comprises similar video shots. Yeo utilizes at least the first frame of a cluster or shot as a representative frame for the entire shot. Yeo terms this first frame as a "key frame." The key frame of Yeo includes both foreground and background information. This is to be expected, since the purpose of the Yeo key frame is simply to represent typical imagery within the scene, and such representation necessarily requires the representation of foreground and background information typical of that scene. A hierarchical graph building technique is employed to provide a graphical means of transitioning between clusters or shots within clusters. In this manner, a browser may identify shots, or clusters of shots, having similar video imagery (e.g., a particular speaker or a particular image). It is crucial to note that the Yeo arrangement is not directed towards a layered representation of video or image information. Rather, the Yeo

arrangement is directed towards the clustering of similar video imagery in a manner allowing rapid retrieval by a video browser utilizing a graphical metaphor to arrange and present the clustered video information.

Zhang et al. discloses a system for automatic video segmentation and key frame extraction for video sequences having both sharp and gradual transitions. Zhang discloses an automatic video content parser for parsing video shots such that they are represented in their native media and retrievable based on their visual content. The Zhang system provides methods for temporal segmentation of video sequences into individual camera shots using a twin comparison method. The method is capable of detecting camera shots defined by sharp breaks and gradual transitions, such as transitions formed using editing techniques such as dissolve, wipe, fade-in and fade-out. Content based key frame selection of individual shots is provided by analyzing the temporal variation of video content and selecting a key frame once the difference of content between the current frame and a preceding selected key frame exceeds a set of preselected thresholds. That is, a key frame according to Zhang comprises the frame following a previous key frame having a content difference exceeding a threshold difference.

ARGUMENT

THE ISSUES UNDER 35 U.S.C. §103

It is submitted that a reasonable interpretation of the references as proposed by the Examiner in the various Office Actions would not have resulted or made obvious the invention recited in the appellants' claims.

A. Whether Claims 1-3, 11 and 21-23 are patentable under 35 U.S.C. §103(a) over Adelson (U.S. Patent No. 5,706,417, issued January 6, 1999) in view of Yeo et al. (U.S. Patent No. 5,821,945, issued October 13, 1998) and Shibata et al. (Content-Based Structuring of Video Information, 0-8186-7436-9/96, 1996 I.E.E.E.).

Claims 1-3, 11 and 21-23 stand rejected by the Examiner (per comment 1 of the final Office Action) under 35 U.S.C. §103(a) as being unpatentable over the Adelson patent (U.S. patent No. 5,706,417, issued January 6, 1998) in view of the Yeo, et al. patent (U.S. patent No. 5,821,945, issued October 13, 1998) and the Shibata et al. paper (Content-Based Structuring of Video Information, 0-8186-7436-9/96, 1996 I.E.E.E.). The Appellants respectfully traverse.

In the Advisory Action, the Examiner stated that:

> "[Adelson] teaches layered representation for image coating wherein all but the dynamic object in the foreground may be joined to form a background image mosaic, and, for example the ball traveling in the foreground associates a plurality of foreground images with a background image. Yeo teaches key frame as the first frame occurring in a segment, similar to the background mosaic image being the first frame in the sequence." (emphasis added by Appellants)

The Appellants strongly disagree with the Examiner's characterization of both the Adelson reference and the Yeo reference. As will be discussed in more detail below (and per the reference summaries above), there is absolutely no teaching within Adelson of joining background layers of image frames to form a background image mosaic. Furthermore, the key frame taught by Yeo has absolutely nothing to do with the key frame of the present invention. The Yeo key frame is simply the first frame of a shot comprising a plurality of frames. In fact, Yeo provides (for the Abstract of the Invention) that "video shots are first identified and a collection of key frames is used to represent each video segment." Thus, Yeo utilizes a plurality of frames to represent a shot or segment, where each of the frames comprises a standard video frame including all background and foreground information within that frame. The Yeo key frame includes information from only one frame, rather than the mosaic information of the claimed key frame of the present invention. The use of similar terminology (i.e., "key frame") does not necessarily mean that concepts so termed are the same. As will be discussed in more detail below, the Examiner's mischaracterizations of both the Adelson and Yeo references, upon which all the Examiner's rejections are based, requires that the Board reject the Examiner's contentions with respect to the claims under appeal.

Inherency

As noted in M.P.E.P. §2112, the Examiner must provide a rationale or evidence tending to show inherency. As noted in *Ex Parte Levy*, 17 U.S.P.Q. 2d 1461, 1464 (Board of Patent Appeals & Interferences 1990), "In relying on the theory of inherency, the

Examiner must provide a basis in fact and/or technical reasoning to reasonably support the determination that the allegedly inherent characteristic necessarily flows form the teachings of the applied prior art."

The Examiner contends that Adelson inherently teaches forming a background mosaic image (an image found by combining background layers of a plurality of images). This is simply not true, since there is no need within the Adelson arrangement to form such an image to accomplish the purposes of Adelson, and certainly no disclosure of such processing. As discussed in Adelson, blurred portions proximate a foreground object are undesirable. Within a single image frame, a layer occluded by an object layer may be processed according to, for example, a blur map to address the blurring problem. There is absolutely no teaching, nor is there any need, for a mosaicing of a plurality of background layers within a scene to effect this solution. The Adelson reference discloses a frame-by-frame method of processing. There is no intra-scene or multiple frame mosaicing required and, therefore, it cannot be said that any mosaicing is inherent to Adelson.

The Examiner contends that Yeo inherently teaches mosaicing of background layers to form a key frame. This is simply not true, since there is no need for Yeo to perform such mosaicing to accomplish the purposes of Yeo and certainly no disclosure of such mosaicing. The Yeo arrangement simply utilizes, for example, a first frame of a shot as a key frame. The purpose of the Yeo key frame is to provide a representative image useful in categorizing the entire shot or scene. Thus, the Yeo key frame cannot simply be a background image; rather, the Yeo key frame must include background and foreground imagery such that the shot or scene is appropriately presented for a subsequent viewer depending on the

key frame to categorize the shot. That is, if the key frame is to represent the imagery associated with a shot or scene, then the key frame must include representative background and representative foreground imagery. In stark contrast, the claimed key frame of the present invention comprises background information and, more particularly, the key frame comprises a mosaic of background information from each of a plurality of frames forming the scene. This is entirely unlike than the Yeo arrangement. It simply cannot be the case that the claimed invention is inherently disclosed in Yeo since the Yeo key frame concept teaches away from the claimed key frame and the claimed key frame cannot in any way be construed as being necessary to achieve the purposes of the Yeo arrangement.

The Examiner's characterization of the Yeo arrangement is inaccurate and extremely misleading. Moreover, even this extremely stretched characterization of the Yeo arrangement would still fail to bridge the gap between the Adelson reference and the claimed invention. Therefore, the combination of Adelson and Yeo does not teach scene background layer mosaic representation, let alone the particular representation used to generate the key frame as claimed.

Impermissible Use of References

As stated by the Federal Circuit (*In re Fritch*, 972 F.2d 1260, Fed. Cir. 1992:

> "It is impermissible to use the claimed invention as an instruction manual or 'template' to piece together the teachings of the prior art so that the claimed invention is rendered obvious. This court has previously stated that 'one cannot use hindsight reconstruction to pick and choose among isolated disclosures in the prior art

> to deprecate the claimed invention.' (citing *In re Fine*, 837, F.2d 1071, Fed. Cir. 1988)."

It is respectfully submitted that the Examiner has used the claimed invention as a template to piece together the teachings of several patents in an attempt to reconstruct the claimed invention. Moreover, it is submitted that the teachings of the various patents may not be operably combined, since, as will be explained in more detail below, the cited references disclose disparate technologies. Finally, even if the cited references could somehow be operably combined, the resulting combination would still fail to disclose or suggest the claimed inventions.

Improper Combination of References

"Obviousness can only be established by combining or modifying the teachings of the prior art to produce the claimed invention where there is some teaching, suggestion, or motivation to do so found either in the references themselves or in the knowledge generally available to one of ordinary skill in the art." (*In re Fine*, 837 F.2d 1071, Fed. Cir. 1988).

The Appellants fail to understand the teaching, suggestion or motivation to initially combine the Adelson, Yeo and Shibata references, irrespective of whether these references in combination do suggest the claimed method (they do not). Specifically, the Examiner seems to assume, *a priori*, that Adelson, Yeo and Shibata may be operably combined without any explanation whatsoever. The Appellants continue to respectfully disagree with his position. As will be discussed in more detail below, the text-based representation of video information provided by Shibata for the benefit of film directors and the like needing a textual description of, for example, a film cannot be operably

combined with the layered image representation of Adelson and the content-based video browsing method and apparatus of Yeo.

References Fail to Suggest the Invention

The Examiner attempts to bridge the substantial gaps between the Adelson and Yeo arrangements, either singly or in combination, and the present invention using the Shibata patent. It is respectfully submitted that the Shibata reference cannot be operably combined with either the Adelson or the Yeo arrangement. Moreover, even if the Shibata reference could be operably combined with either of these arrangements (or both), the resulting combination would still fail to disclose or suggest the claimed invention.

The Appellants urge the Board to categorically reject the Examiner's use of hindsight, assertions of implicit teachings and mischaracterizations of the cited references to arrive at the untenable positions clearly evident in the prosecution history and referenced in this Appeal Brief.

Adelson, Yeo and Shibata, either singly or in combination, fail to disclose or suggest the invention per amended claim 1, which reads as follows (labels inserted to simplify the discussion):

> "A method for comprehensively representing video information in a manner facilitating indexing of the video information, comprising the step of:
>
> (a) segmenting a video stream into a plurality of scenes, each of said scenes comprising at least one video frame;

(b) dividing, using intra-scene motion analysis, at least one of said plurality of scenes into at least one scene foreground layer and a scene background layer;

(c) representing each scene background layer as a mosaic, said background layer mosaic defining a key frame of a respective scene; and

(d) representing each of said at least one video frames forming said scene as a difference between initial video frame imagery and a respective portion of said key frame."

With respect to step (b), Adelson utilizes intra-frame motion to define layers within a video frame. By contrast, the subject invention claims "intra-scene motion analysis" to divide a "scene[] into at least one scene foreground layer and a scene background layer." Intra-scene analysis utilizes a plurality of frames within a scene, not just a single frame, to divide a scene (not a frame) into layers. Thus, Adelson provides a different structure, operating in a different manner to achieve a different purpose than the claimed invention. To the extent that Adelson teaches defining layers, there is absolutely no teaching within Adelson of combining or mosaicing layers from a plurality of image frames within a scene to form a combined or mosaiced background layer.

The teachings of Yeo do not bridge the considerable gap between Adelson and the claimed invention. Specifically, assuming arguendo that Adelson and Yeo were to be somehow operatively combined, the resulting combination would still lack the claimed element (b). That is, the resulting combination would, at most, provide for segmenting a video stream and processing individual

frames within a scene to provide a series of frame-specific layers.

With respect to step (c), there is absolutely no teaching of forming a mosaic within either of the two references, much less the claimed step of "representing each scene background layer as a mosaic," where the "background layer mosaic defin[es] a key frame of a respective scene."

The Examiner contends that Adelson "teaches combining the foreground and background images to produce a video image, thereby implicitly teaching mosaic representation (column 2, lines 15-21; column 6, lines 50-55)." The Appellant respectfully disagrees, for at least the following reasons.

The portions of Adelson cited by the Examiner (page 2 and 6) notes that multiple motion vectors exist where, for example, the edge of a moving foreground object is blurry. Such motion blur and/or focus blur occurs in the case of, for example, an object such as a baseball moving rapidly across a display or viewing window. Simply put, the portions of text cited by the Examiner address intra-frame motion of a foreground object and the effect of the motion of that object on the clarity of traversed background imagery. This is entirely unlike the claimed invention, in which intra-scene (i.e., within a scene formed using a plurality of frames) processing is utilized to provide image layering for subsequent use in a mosaic representation.

To clarify elements (c) and (d) of claim 1, the Board is referred to FIG. 7 and the associated text beginning on page 20 of the subject application. Specifically, the graphical representation depicted in FIG. 7 is of a boat sailing from right to left. In the right most background scene 740, a sun 744 and clouds 746 are found. In the left most background scene 730, a

remainder portion of the clouds 736 and a dock 739 are found. The invention utilizes mosaic technology to combine the background images to produce a unified background image 760 including the dock 769, a cloud 766, and the sun 764. This background image is used as a first frame in a sequence of frames depicting the scene of the boat sailing. The remaining frames F_1 through F_m of the scene 750 incorporates primarily foreground imagery of the boat moving within the frame.

It is important to note that there is absolutely no teaching in the Adelson reference of combining background imagery from different frames to form a key or anchor background image which is then associated with a plurality of foreground images such as depicted in FIG. 7. The portion of text cited by the Examiner only supports the notion that a foreground object in motion tends to distort or blur background imagery proximate the object in motion. There is absolutely no teaching or suggestion within the Adelson reference that in the mosaic technique is employed in the manner described and claimed in claim 1 of the subject invention.

The Appellant submits that it cannot be reasonably argued that combining a foreground and background image implicitly teaches a mosaic representation. This is because the mosaic representation of claim 1 comprises the combining of at least portions of multiple images, not of multiple layers within a single image (as provided by Adelson).

The Examiner contends that Yeo teaches the "key frame" limitation of claim 1. The Appellant respectfully disagrees, for at least the following reasons.

It is noted that the Yeo arrangement utilizes scene cut detection to segment a video screen into a plurality of shots or scenes. However, the claimed invention is not simply the dividing

of a video stream into a plurality of scenes. Rather, the subject invention of claim 1 comprises a plurality of steps including the step of segmenting a video stream into a plurality of stream and, additionally, processing that step using various processing steps not shown in either of the Yeo arrangement or the Adelson arrangement.

The Examiner contends that Shibata "teaches segmenting a video sequence, with individual video frames being the smallest unit of any segment [and] the use of a basic segment which is a collection of video frames having the same vector expressions, assuming a collection of basic segments as the initial layer, and creating new layers by adding a segment to the previously processed layer, thus teaching a method for providing background mosaic, and intra-scene motion analysis." The Appellants strongly disagree.

As with Adelson and Yeo, the Examiner has misconstrued the teachings of Shibata. For example, the collection of "basic segments" forming an initial layer and the creation of new layers by averaging segment pairs in previous layers is construed by the Examiner as teaching a method for providing a background mosaic and for teaching intra-scene motion analysis. This is simply not the case, as will be discussed below.

Shibata teaches content-based structuring of video information using textual descriptions. It is noted that Shibata defines (per Section 3.1) "video structuring" as an operation which divides a video sequence into "segments" and describes the hierarchical relations between the segments. It is also noted that the description in Shibata of the relations between segments is a textual description intended to provide a human readable description of the underlying video scene such that the underlying video may be readily processed within the context of a video editing environment or studio environment, e.g. by a director. Specifically, a descriptive component (DC) is defined by Shibata as key words or elemental words that constitute short sentences

which may be divided into several groups (see Section 2.). With respect to video structuring, the categories of visual objects, actions of the object, and state of the object are used. The descriptive components (DCs) are mapped (see FIG. 1) as a script which indicates the presence or absence of particular descriptive components within the video sequence in time.

The "vector expressions" of Shibata are not motion vectors. Rather (per section 3.1), the Shibata "vector expressions" are merely representations of the duration of descriptive components in terms of time or segment length. The Shibata "vector expressions" should not be equated with the motion vectors discussed in the instant patent application.

It can be seen in FIG. 2 of Shibata that each "layer" is formed by averaging "basic segments" of a lower layer. That is, as depicted in FIG. 2, where M basic segments are provided, the M^{th} layer includes the M "basic segments." By averaging the vector expressions of respective adjacent basic segments within the M^{th} layer, an M^{th} minus 1 layer is formed which includes M/2 basic segments. Each of the M/2 basic segments comprises the averaged vector expressions of the two basic segments within the M^{th} layer. Similarly, for each succeeding layer, respective pairs of basic segments or derived (i.e., averaged) basic segments are themselves averaged to produce the next layer. A top layer or M = 1 layer comprises the average of all of the vector expressions of the basic segments forming the M^{th} layer. This averaging or decimation of information on a layer-by-layer basis cannot be construed to teach the intra-scene layering of the subject invention. In fact, it is impossible to reconstruct lower layer video information using the upper layer information.

Since the references, either singly or in combination, do not disclose or suggest the claimed invention it is respectfully submitted that the invention of claim 1, at least as amended, is patentable over the cited references. Moreover, since independent claim 21 includes limitations similar to those found in

independent claim 1, it is submitted that claim 21 is patentable for at least the reasons discussed above with respect to claim 1. Therefore, the Appellant submits that claims 1, 17 and 21, as they now stand, fully satisfy the requirements of 35 U.S.C. §103 and are patentable thereunder.

Furthermore, all of the remaining dependent claims depend, either directly or indirectly, from claims 1 or 21 and recite additional features therefor. As such and for the exact same reasons set forth above, the Appellants submit that none of these claims is obvious with respect to the teachings of the cited references. Therefore, the Appellant submits that all these dependent claims also fully satisfy the requirements of 35 U.S.C. §103 and are patentable thereunder.

Further with respect to claim 2, the cited references fail to disclose or suggest at least the limitation of "computing, for at least one of said scene foreground and background layers, one or more content-related appearance attributes." The claimed "appearance attributes," are defined on page 17, lines 6-14 of the specification as follows:

> "Appearance attributes of each representative frame and each object within a scene are computed independently and associated with the scene for subsequent indexing and retrieval of, e.g., the stored video. The appearance attributes consist of color and texture distributions, shape descriptions, and compact representation in terms of outputs of multiple scale, multiple orientation and multiple moment Gaussian and Gabor-like filters. These attributes are organized in terms of data structures that will allow similarity queries to be answered very efficiently. For example, multi-dimensional R-tree data structures can be used for this purpose."

In addition, as noted on page 16, lines 31 through 33 of the application, "appearance attributes … are computed only for

"representative frames," e.g. mosaics or key frames within a scene." Thus, the step of computing within claim 2 does not generate appearance attributes for each and every frame of a video sequence as disclosed in Adelson.

Thus, unlike the present invention of claim 2, and regardless of whether the maps of Adelson may be fairly characterized as appearance attributes, the various maps of Adelson are provided for each layer of each and every frame of a video sequence.

Therefore, it is respectfully submitted that claim 2 is patentable over the cited references for the above additional reasons. Moreover, since claim 22 includes limitations similar to those found in claim 2, it is respectfully submitted that claim 22 is also patentable over the cited references for at least the reasons discussed above with respect to claim 2. Furthermore, since claims 3, 5-10 and 23-26 depend, either directly or indirectly from claim 2 or 22 and recite additional features thereto, it is respectfully submitted that these claims are also patentable for at least the reasons discussed above with respect to claim 2.

B. Whether Claims 4 and 24 are patentable Under 35 U.S.C. §103(a) Over Adelson (U.S. Patent No. 5,706,417 issued January 6, 1999) in View of Yeo et al. (U.S. Patent No. 5,821,945, issued October 13, 1998) and Shibata et al. (Content-Based Structuring of Video Information, 0-8186-7436-9/96, 1996 I.E.E.E.) as applied to Claims 1 and 22, respectively, and further in view of Jaillon et al. (Image Mosaicing Applied To Three-Dimensional Surfaces: 1051-4651/94-1994 I.E.E.E.).

The Examiner has rejected claims 4 and 24 as being obvious per the Adelson patent in view of the Yeo patent and Shibata paper as applied to claims 1 and 22, and further in view of the Jaillon paper. This rejection is respectfully traversed. The Appellants

contend that claims 4 and 24 are patentable for at least the reasons discussed above with respect to claims 2 and 22, from which they respectively depend.

The Appeals Board is respectfully directed to the above discussion of the references. As previously noted, Appellants respectfully submit that the Examiner has severely misconstrued these references to arrive at a logically untenable position with respect to the base claims.

Specifically, there is absolutely no teaching in the references of combining background imagery from different frames to form a key-frame or anchor background image which is then associated with a plurality of foreground images, such as depicted in FIG. 7 of the subject application.

As noted by the Examiner in comment 1 of the final Office Action: "Adelson does not teach segmenting a video stream into scenes, and scenes into frames including a key frame, and the use of intra-scene motion analysis." The Appellants thank the Examiner for noting this important distinction.

The Jaillon reference discloses a method of mosaicing images lying on a three-dimensional surface by modeling the three-dimensional surface as a two-dimensional surface, merging the images using a two-dimensional mosaicing technique, applying corrections to the resulting mosaic and returning the corrected mosaic to three-dimensional space. The Jaillon arrangement is primarily directed to solving distortions and other problems associated with still image mapping between two and three-dimensional surfaces.

The Adelson, Yeo, Shibata and Jaillon arrangements, either singly or in combination, fail to disclose or suggest the invention of claims 4 and 24 in which "said mosaic representation

comprises one of a two-dimensional mosaic, a three-dimensional mosaic, and a network of mosaics."

In contrast to the language of claims 4 and 24, there is no teaching in any of the references of a network of mosaics. Moreover, there is no teaching in the references of the video processing steps including inter-scene processing steps of the respective base claims 1 and 22. It is noted that the claimed mosaic representation of the subject invention comprises a mosaic of background imagery for a plurality of image frames. Nothing within the Jaillon reference, or the other references, teaches such a mosaic.

Therefore, it is respectfully submitted that the references, either singly or in combination, do not disclose or suggest the invention of claims 4 and 24. As such, the Appellants submit that claims 4 and 24 fully satisfy the requirements of 35 U.S.C. §103 and are patentable thereunder.

C. Whether Claims 5-8 are Patentable Over Adelson (U.S. Patent No. 5,706,417, issued January 6, 1999), Yeo et al. ((U.S. Patent No. 5,821,945, issued October 13, 1998) and Shibata et al. (Content-Based Structuring of Video Information, 0-8186-7436-9/96, 1996 I.E.E.E.) as applied to Claim 2 and further in view of Jaillon et al. (Image Mosaicing Applied To Three-Dimensional Surfaces: 1051-4651/94-1994 I.E.E.E).

The Examiner has rejected claims 5-8 over the Adelson patent, Yeo patent and Shibata paper as applied to claim 2, and further in view of the Jaillon paper. This rejection is respectfully traversed. The Appellants contend that claims 5-8 are patentable for at least the reasons discussed above with respect to claim 2, from which they depend either directly or indirectly.

The Appeals Board is respectfully directed to the above discussion of the Adelson, Yeo, Shibata and Jaillon references.

As previously noted, Appellants respectfully submit that the Examiner has severely misconstrued several of these references to arrive at a logically untenable position with respect to the base claims.

The references, either singly or in combination, fail to disclose or suggest the invention of claim 5, as follows:

> "The method of claim 2, wherein said step of computing a content-based appearance attribute for a layer of a scene comprises the steps of:
>
> generating an image pyramid of said layer;
>
> filtering, using one or more filters associated with said content-based appearance attribute, each subband of said image pyramid to produce respective one or more feature maps associated with each subband; and
>
> integrating said one or more feature maps associated with each respective subband to produce respective attribute pyramid subbands, wherein each of said attribute pyramid subbands comprises a content-based appearance attribute subband associated with a corresponding image pyramid subband."

In contrast to the above-quoted claim language, the references fail to disclose or suggest the above steps for computing a content-based appearance attribute for a layer of a scene. The Examiner cites column 1, lines 20-24, of Adelson to disclose the use of sub-bands to encode images. The Appellants agree that sub-band coding for encoding images is known. However, the method of claim 5 is not simply the use of sub-band coding; rather, the method utilizes content-based appearance attributes to perform sub-band filtering to produce respective feature maps which are then integrated to form respective attribute pyramid sub-bands. This is entirely different than the prior art. Moreover, the sub-band coding noted in the Adelson reference is simply one of a plurality of coding techniques purportedly inferior to the techniques provided in Adelson. That is, Adelson

teaches alternative techniques to sub-band coding and, therefore, teaches away from any coding techniques using sub-band coding. Adelson is directed to coding an image represented as a plurality of layers, where each layer has associated with it a respective plurality of maps. Adelson is not directed to sub-band encoding, and certainly not the image pyramid generation and processing steps of claim 5.

As noted by the Examiner, "Adelson and Yeo fail to teach image pyramids." The Examiner contends that "Jaillon teaches the use of image pyramid framework in the alignment process, and converting the input image and the mosaic into Laplacian image pyramids, and applying the alignment to all levels within the respective pyramids." The Examiner then contends that those skilled in the art would use the Jaillon image pyramid in each layer of, presumably Adelson, to achieve better alignment and reproduction of the image pyramid. The Appellants respectfully disagree.

While Jaillon does utilize Laplacian pyramids within the context of fusing still imagery to form a two-dimensional mosaic, it is noted that, per page 256, first column, section 4.2, third paragraph, Jaillon notes that "the height [of the Laplacian pyramids] must be chosen depending on radiometric discontinuities." This limitation in the use of Laplacian pyramids renders the Jaillon teachings of pyramid representation and utilization inappropriate to the teachings of Adelson and Yeo (i.e., any combination of Jaillon, Adelson and Yeo is inappropriate) and, moreover, inappropriate to the operation of the invention as claimed.

With respect to claim 8, the Examiner contends, using an enormous stretch of logic, that the visual cues utilized in Yeo

can somehow be compared to the image pyramids of the present invention and, more specifically, the attribute pyramids of claim 8. This is simply wrong. The hierarchy discussed in Yeo is entirely directed to an organization of video material such that a viewer may find desired imagery using a hierarchical search technique. By contrast, claim 8 is directed to a processing step for producing content-based appearance attributes using sub-band pyramid information.

The Appeals Board is respectfully directed to the abject failure of the Examiner to reasonably characterize the Shibata reference, Yeo reference and Adelson references and the logical deficiencies that flow from such mischaracterization.

Further with respect to claim 6, the Examiner notes that "Adelson discloses the use of intensity map, depth map, blur map [and contrast] change map. Clearly, there is no indication in the Adelson reference of a map corresponding to chrominance or texture attributes. Moreover, to the extent that the named maps implicate luminance attributes, it is noted that the maps, per column 2, lines 55-59, an intensity map "essentially defining the image comprising that layer at a fixed instant in time, e.g., the initial frame of the sequence."

Since the references, either singly or in combination, do not disclose or suggest the invention of claim 5, it is respectfully submitted that the invention of claim 5 is patentable over the cited references. Moreover, since claims 6-8 depend from claim 5, and include additional limitations thereto, it is respectfully submitted that these claims also fully satisfy the requirements of 35 U.S.C. §103 are patentable thereunder.

D. Whether Claims 9 and 10 are patentable under 35 U.S.C. §103(a) over Adelson(U.S. Patent No. 5,706,417, issued January 6, 1999) in View of Yeo et al. (U.S. Patent No. 5,821,945, issued October 13, 1998) and Shibata et al. (Content-Based Structuring of Video Information, 0-8186-7436-9/96, 1996 I.E.E.E.) as applied to Claim 2 and further in view of Barber et al. (U.S. Patent No. 5,751,286, issued May 12, 1998).

The Examiner has rejected claims 9 and 10 as being obvious per the Adelson patent, Yeo patent and Shibata patent as applied to claim 2 and further in view of the Barber patent. This rejection is respectfully traversed.

The Appellants contend that claims 9 and 10 are patentable for at least the reasons discussed above with respect to claim 2, from which claims 9 and 10 depend either directly or indirectly.

The Appeals Board is respectfully directed to the above discussion of the Adelson, Yeo and Shibata references. As previously noted, Appellants respectfully submit that the Examiner has severely misconstrued several of these references to arrive at a logically untenable position with respect to at least the base claims.

Barber discloses an image query system and method wherein images in an image data base are searched in response to queries which include the visual characteristics of the images such as colors, textures, shapes, and sizes as well as by textual tags appended to the images.

It is respectfully submitted that the Barber reference fails to bridge the substantial gap between the previously cited references and the invention of claims 9 and 10. Therefore, it is respectfully submitted that claims 9 and 10 are patentable over the cited references including the Barber reference.

E. Whether Claims 13-14 are patentable over Adelson (U.S. Patent No. 5,706,417, issued January 6, 1999) in view of Yeo et al. (U.S. Patent No. 5,821,945, issued October 13, 1998) and

Shibata et al. (Content-Based Structuring of Video Information, 0-8186-7436-9/96, 1996 I.E.E.E.) as Applied to Claim 1,and further in view of Zhang (U.S. Patent No. 5,635,982, issued June 3, 1997).

The Examiner has rejected claims 13-14 as being obvious per the Adelson patent, Yeo patent and Shibata paper as applied to claim 1 and further in view of the Zhang patent. This rejection is respectfully traversed.

The Appellants contend that claims 13-14 are patentable for at least the reasons discussed above with respect to claim 1, from which they depend either directly or indirectly.

The Appeals Board is respectfully directed to the above discussion of the Adelson, Yeo and Shibata references. As previously noted, Appellants respectfully submit that the Examiner has severely misconstrued several of these references to arrive at a logically untenable position with respect to at least the base claims.

Zhang et al. discloses a system for automatic video segmentation and key frame extraction for video sequences having both sharp and gradual transitions. Zhang discloses an automatic video content parser for parsing video shots such that they are represented in their native media and retrievable based on their visual content. The Zhang system provides methods for temporal segmentation of video sequences into individual camera shots using a twin comparison method. The method is capable of detecting camera shots defined by sharp breaks and gradual transitions, such as transitions formed using editing techniques such as dissolve, wipe, fade-in and fade-out. Content based key frame selection of individual shots is provided by analyzing the temporal variation of video content and selecting a key frame once the difference of content between the current frame and a preceding selected key frame exceeds a set of preselected thresholds. That is, a key frame according to Zhang comprises the frame following a previous

key frame having a content difference exceeding a threshold difference.

In addition to the limitations of claim 1, claim 13 specifically defines the step of segmenting of claim 1 as "generating a descriptor vector...; calculating a difference between descriptor vectors of successive frames; and generating a scene cut indicium in response to said calculated difference exceeding a threshold level." Such teaching is not met by the Zhang reference.

It is respectfully submitted that the Zhang patent fails to bridge the substantial gap between the previously cited references and the invention of claims 13 and 14. Therefore, it is respectfully submitted that claims 13 and 14 are patentable.

F. Whether Claims 17-20 are patentable under 35 U.S.C. §103(a) over Barber et al. (U.S. Patent No. 5,751,286, issued May 12, 1998) in view of Yeo et al. (U.S. Patent No. 5,821,945, issued October 13, 1998) in view of Shibata et al. (Content-Based Structuring of Video Information, 0-8186-7436-9/96, 1996 I.E.E.E.).

The Examiner has rejected claims 17-20 as being obvious per the Adelson patent in view of the Barber patent, the Yeo patent and Shibata paper. This rejection is respectfully traversed.

The Appellants contend that claims 17-20 are patentable for at least the reasons discussed above with respect to claim 1 (which is similar in scope to claim 20).

The Appeals Board is respectfully directed to the above discussion of the Barber, Yeo and Shibata references. As previously noted, the Examiner has severely misconstrued at least the Yeo and Shibata references to arrive at a logically untenable position with respect to at least the base claims.

G. Whether Claims 25-26 are patentable over Adelson (U.S. Patent No. 5,706,417, issued January 6, 1999) in view of Yeo et al. (U.S. Patent No. 5,821,945, issued October 13, 1998) and Shibata et al. (Content-Based Structuring of Video Information, 0-8186-7436-9/96, 1996 I.E.E.E.) as applied to Claim 22 and further in view of Jaillon et al. (Image Mosaicing Applied To Three-Dimensional Surfaces: 1051-4651/94-1994 I.E.E.E).

The Examiner has rejected claims 25-26 as being obvious per the Adelson patent in view of the Yeo patent and Shibata paper as applied to claim 22, and further in view of the Jaillon paper. This rejection is respectfully traversed.

The Appellants contend that claims 25-26 are patentable for at least the reasons discussed above with respect to claim 22, from which claims 25-26 depend, either directly or indirectly.

The Appeals Board is respectfully directed to the above discussion of the Adelson, Yeo, Shibata and Jaillon references. As previously noted, Appellants respectfully submit that the Examiner has severely misconstrued several of these references to arrive at a logically untenable position with respect to at least the base claims. These references are clearly deficient in terms of independent claim 22, and also in terms of dependent claims 25 and 26.

CONCLUSION

For the extensive reasons advanced above, appellants respectfully but forcefully content that each claim is patentable. Therefore, reversal of all rejections is courteously solicited.

Respectfully submitted,

8/28/00

Eamon J. Wall, Attorney
Reg. No. 39,414
(732) 530-9404

Thomason, Moser & Patterson LLP
Attorneys at Law
595 Shrewsbury Avenue, 1st Floor
Shrewsbury, New Jersey 07702

CLAIMS

1. A method for comprehensively representing video information in a manner facilitating indexing of the video information, comprising the step of:

segmenting a video stream into a plurality of scenes, each of said scenes comprising at least one video frame;

dividing, using intra-scene motion analysis, at least one of said plurality of scenes into at least one scene foreground layer and a scene background layer;

representing each scene background layer as a mosaic, said background layer mosaic defining a key frame of a respective scene; and

representing each of said at least one video frames forming said scene as a difference between initial video frame imagery and a respective portion of said key frame.

2. The method of claim 1, further comprising the steps of:

computing, for at least one of said scene foreground and background layers, one or more content-related appearance attributes; and

storing, in a database, said scene content-related appearance attributes or said mosaic representations.

3. The method of claim 2, further comprising the steps of

storing representations of said plurality of scenes in a mass storage unit; and

retrieving, in response to a database query, scenes associated with content-related appearance attributes defined in said database query.

4. The method of claim 1, wherein said mosaic representation comprises one of a two dimensional mosaic, a three dimensional mosaic and a network of mosaics.

5. The method of claim 2, wherein said step of computing a content-based appearance attribute for a layer of a scene comprises the steps of:

generating an image pyramid of said layer;

filtering, using one or more filters associated with said content-based appearance attribute, each subband of said image pyramid to produce respective one or more feature maps associated with each subband; and

integrating said one or more feature maps associated with each respective subband to produce respective attribute pyramid subbands, wherein each of said attribute pyramid subbands comprises a content-based appearance attribute subband associated with a corresponding image pyramid subband.

6. The method of claim 5, wherein said content-based appearance attribute comprises at least one of a luminance attribute, a chrominance attribute and a texture attribute.

7. The method of claim 5, wherein said step of filtering further comprises the step of:

rectifying each of said one or more feature maps associated with each subband.

8. The method of claim 5, further comprising the step of:

collapsing said attribute pyramid subbands to produce a content-based appearance attribute.

9. The method of claim 2, further comprising the step of:

receiving a request for video information substantially matching a desired content-related appearance attribute; and

retrieving video frames or scenes having at least one layer associated with content-related appearance attributes substantially matching said desired content-related appearance attribute.

10. The method of claim 9, wherein said step of receiving a request comprises the steps of:

identifying a query type and a query specification, said query type comprising one of a luminance, chrominance and texture query type, said query specification defining a desired property of said identified query type;

selecting a predetermined filter type associated with said identified query type; and

calculating, using said predetermined filter type and said desired property, a desired content-related appearance attribute, said desired content-related appearance attribute being suitable for comparing to said content-related appearance attributes stored in said database.

11. The method of claim 1, further comprising the steps of:

storing, in a database, ancillary information associated with one or more layers or frames of one or more scenes.

13. The method of claim 1, wherein said step of segmenting comprises the steps of:

generating a descriptor vector of a predetermined type for each video frame of a video information stream;

calculating a difference between descriptor vectors of successive frames; and

generating a scene cut indicium in response to said calculated difference exceeding a threshold level.

14. The method of claim 1, wherein said step of segmenting comprises the steps of:

generating, in a first pass, a descriptor vector of a predetermined type for each video frame of a video information stream;

calculating, using said generated descriptor vectors, a descriptor vector threshold level;

calculating, in a second pass, a difference between descriptor vectors of successive frames; and

generating a scene cut indicium in response to said calculated difference exceeding a threshold level.

17. A method for browsing a video program stored in a mass storage unit, said video program comprising a plurality of scenes, said scenes comprising a plurality of video frames including a key frame comprising a mosaic of an intra-scene background layer, said method comprising the steps of:

providing a database associated with the stored video program, said database comprising attribute information associated with at least a representative portion of said plurality of video frames forming each scene;

formulating a query utilizing attribute information associated with a desired video frame;

searching said database to identify video frames substantially satisfying said query; and

retrieving, from said mass storage unit, one or more of said identified video frames.

18. The method of claim 17, wherein said step of formulating a query comprises the steps of:

selecting a query type;

selecting a query specification; and

computing a multi-dimensional feature vector using said query type and query specification.

19. The method of claim 18, wherein said query specification is selected by identifying a portion of a displayed image, and said multi-dimensional feature vector is calculated using said query type and said identified potion of said displayed image.

20. The method of claim 19, further comprising the steps of:

formatting, for subsequent presentation on a display device, each scene including one or more of said identified video frames; and

transmitting said formatted scenes.

21. A computer-readable medium having stored thereon a plurality of instructions, the plurality of instructions including instructions which, when executed by a processor, cause the processor to perform the steps of:

(a) segmenting a video stream into a plurality of [video]scenes, each of said scenes comprising at least one video frame;

(b) dividing, using intra-scene motion analysis, at least one of said plurality of scenes into at least one scene foreground layer and a scene background layer;

representing each scene background layer as a mosaic, said background layer mosaic defining a key frame of a respective scene; and

representing each of said at least one video frames forming said scene as a difference between initial video frame imagery and a respective portion of said key frame.

22. The computer-readable medium of claim 21, further having stored thereon an additional plurality of instructions, the additional plurality of instructions including instructions which, when executed by a processor, cause the processor to perform the additional steps of:

computing, for at least one of said scene foreground and background layers, one or more content-related appearance attributes; and

storing, in a database, said scene content-related appearance attributes or said mosaic representations.

23. The computer-readable medium of claim 22, further having stored thereon an additional plurality of instructions, the additional plurality of instructions including instructions which, when executed by a processor, cause the processor to perform the additional steps of:

storing representations of said plurality of scenes in a mass storage unit; and

retrieving, in response to a database query, scenes associated with content-related appearance attributes defined in said database query.

24. The computer-readable medium of claim 22, wherein said mosaic representation comprises one of a two dimensional mosaic, a three dimensional mosaic and a network of mosaics.

25. The computer-readable medium of claim 22, wherein the stored instruction of computing a content-based appearance attribute for a layer of a scene, when executed by a processor, cause the processor to perform the steps of:

generating an image pyramid of said layer;

filtering, using one or more filters associated with said content-based appearance attribute, each subband of said image pyramid to produce respective one or more feature maps associated with each subband; and

integrating said one or more feature maps associated with each respective subband to produce respective attribute pyramid subbands, wherein each of said attribute pyramid subbands comprises a content-based appearance attribute subband associated with a corresponding image pyramid subband.

26. The computer-readable medium of claim 25, wherein said content-based appearance attribute comprises at least one of a luminance attribute, a chrominance attribute and a texture attribute.